



ARTICLE

Rousseau and Kant on the Moral Value of Compassion

David James

University of Warwick, Coventry, UK
Email: d.n.james@warwick.ac.uk

Abstract

Despite Rousseau's acknowledged influence on Kant, the moral value of compassion (or pity) is regarded as a major difference between their theories of morality. Pity plays a fundamental role in Rousseau's theory of moral relations, whereas Kant appears suspicious of compassion. I argue that Kant nevertheless accords compassion a significant moral value, not only because it provides an appropriate supplementary incentive when the incentive of duty is not sufficient to motivate action but also because of the role it plays in attuning individuals to the moral status of others. Rousseau's account of pity in *Emile* helps to explain how compassion can play this role.

Keywords: compassion; dignity; moral development; pity; Rousseau; sympathy

1. Introduction

Kant acknowledged the decisive influence of Rousseau's writings on his own moral outlook (*ROFBS*, 20: 44).¹ It would therefore be surprising if this moral influence did not make its presence felt in Kant's philosophy and if studies that compare Rousseau's and Kant's ideas did not accordingly stress affinities rather than differences.² Yet one important difference has been emphasised: Rousseau's and Kant's different attitudes towards compassion. Arthur Schopenhauer firmly distinguishes Kant's moral philosophy from Rousseau's when he claims that Rousseau 'hit upon the truth and stirred the heart' by making compassion (*Mitleid*) fundamental to his account of moral action, whereas Kant completely denies the moral value of compassion (Schopenhauer 2009: 232). Friedrich Nietzsche places Kant in the camp of those philosophers who have a 'low opinion of pity' (*Geringschätzung des Mitleidens*), as opposed to those philosophers who are guilty of an overestimation (*Überschätzung*) of its value (Nietzsche 1994: Preface, sec. 5). Although Nietzsche does not claim that Kant accords *no* value to compassion, it is implied that any value that it does possess is extremely limited. More recently, Martha Nussbaum has contrasted 'defenders of pity', including Rousseau, with opponents of it that include Kant as well as Nietzsche (Nussbaum 1996: 28, 41–8).

As we shall see, Kant identifies compassion as a type of sympathy that is directed at a specific object, which is the suffering of others. I seek to show not only that this type

© The Author(s), 2024. Published by Cambridge University Press on behalf of Kantian Review. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

of sympathy, contrary to what Schopenhauer, Nietzsche, and Nussbaum claim, plays an important role in Kant's moral theory, despite those passages that suggest that he devalues it, but also that elements of Rousseau's account of pity in his educational treatise *Emile* help to explain this role, and thus the moral value that compassion has for Kant.³ The moral role and value of compassion, or pity, does not, therefore, constitute a fundamental difference between Rousseau's and Kant's moral theories. Rather, for both philosophers, compassion has significant moral value because of its essential role in the moral life of human beings, even if Kant limits the value of compassion in a way that Rousseau does not. The fact that this value concerns the role of compassion in the moral development of a human being is indicated by the following passage from Kant's *The Metaphysics of Morals*:

But while having compassion for others [*Mitleid . . . mit Anderen*], and thus also sharing their joys [*Mitfreude*], is not in itself a duty, active sympathy [*thätige Theilnehmung*] for their fate is one; and to this end it is therefore an indirect duty to cultivate the compassionate [*mitleidige*] natural (aesthetic) feelings in us, and to make use of them as so many means to sympathy based on moral principles and the feeling appropriate to them. – It is therefore a duty not to avoid the places where the poor who lack the most basic necessities are to be found but rather to seek them out, and not to shun sickrooms or debtors' prisons and so forth in order to avoid the painful [feeling of] compassion [*Mitgefühl*] that one may not be able to resist. For this is still one of the impulses that nature has implanted in us to do what the representation of duty alone might not accomplish. (*MM*, 6: 457; translation modified)

Kant, here, presupposes the existence of certain natural feelings in the observer: feelings that encounters with the suffering of others trigger rather than create, though, in the absence of encounters of the relevant kind, these feelings are liable to become dulled or to be repressed to such an extent that they will not be able to perform the function which Kant assigns to them, hence the need to cultivate these feelings. The function in question is to supplement the representation of duty, which, as we shall see, means providing an additional but subordinate incentive when it comes to doing what is morally right or omitting to do what is morally wrong. Thus, the desire to avoid the type of environment mentioned in the passage above can ultimately be explained in terms of wanting to avoid experiencing certain feelings, whereas witnessing the suffering of others ensures that these feelings remain sufficiently lively and powerful within oneself.

Compassion involves sharing the feelings of others but without necessarily experiencing the same suffering as they do. Rather, what matters is that one can imagine what the other person is experiencing based on one's own experiences of suffering and one's awareness of what it means to suffer, and that this way of sharing the feelings of others produces a type of feeling in oneself that directs one to act morally. Feeling is, therefore, a necessary feature of compassion in two ways: (1) the feelings experienced by the person who is assumed to be suffering are recreated in oneself by means of an imaginative act, and (2) sharing the feelings of others in this way produces within oneself feelings whose character and intensity depend on the extent to which one believes another human being is suffering, ranging from a mild

emotional unease to a shudder or some other major psychological disturbance accompanied by certain physical symptoms. The pupil in Rousseau's *Emile* provides an example of the second type of emotional and physical response produced by imagining the suffering of another human being or other sentient being, in that he has 'gut reactions at the sounds of complaints and cries', while 'the sight of blood flowing will make him avert his eyes; the convulsions of a dying animal will cause him an ineffable distress before he knows whence come these new movements within him' (OC, 4: 505; E, 374).⁴

In what follows, I argue that Rousseau provides an explanation of why compassion is especially suited to perform an important function in a human being's moral development. This function in turn helps to explain why Kant claims that it is a duty to cultivate those compassionate feelings that he classes as natural ones. The relevant function involves producing an awareness of the moral status of others, thereby making one receptive to moral constraints and demands that follow from this status. I shall nevertheless indicate an essential difference between the moral value that compassion has for Rousseau and the moral value that it has for Kant. This difference is indicated by Kant's reference to 'impulses that nature has implanted in us to do what the representation of duty alone might not accomplish' in the passage from *The Metaphysics of Morals* quoted above. For it suggests that, for Kant, the emotional impulse provided by compassion will no longer possess any genuine moral value once it is no longer required as an additional incentive for doing what morality demands or omitting to do what it forbids. Compassion might also be thought to cease to have any genuine moral value once it has performed the role that it plays in an individual's moral development. I shall argue that the significance of this difference appears less pronounced, however, if a human standpoint is kept in view.

2. Accommodating compassion within Kant's moral theory

The possibility of accommodating compassion within Kant's moral theory in such a way that this emotion serves to provide an additional incentive to moral action can be demonstrated in connection with his remarks on shared or sympathetic feeling (*Mitgefühl*, *theilnehmende Empfindung*), which involves 'sensible feelings of pleasure or displeasure' (MM, 6: 456). While the feeling of pleasure is caused by another person's joy, the feeling of displeasure is caused by another person's pain or anguish (*Schmerzen*). Compassion is a type of shared or sympathetic feeling whose object is another person's suffering and that itself may consist in feelings of inner discomfort, distress, or unease. Witnessing the suffering of others may also produce feelings of discontent or indignation, especially when the suffering is considered to be preventable and undeserved. Compassion can accordingly be identified with a certain type of displeasure.

Yet sympathy cannot be a matter of mere receptivity if it is to promote 'active and rational beneficence' in accordance with 'the duty of humanity'. This duty concerns the view of the human being 'not merely as a rational being but also as an animal endowed with reason' that has 'the capacity and the will to share in others' feelings' (MM, 6: 456). As we have seen, cultivating the relevant feelings by seeking out encounters with the suffering of others is part of what it means to exercise this capacity to share the feelings of others. It is through this willing cultivation of feelings that we do not

remain in a purely passive state in relation to them.⁵ We have also seen that the reason Kant gives for the duty to cultivate the relevant feelings implies that reason by itself tends not to be sufficient to motivate moral action and that an additional incentive is, therefore, needed. Explaining this claim requires taking a closer look at the relation between the incentive provided by compassion and the incentive provided by reason.

On the one hand, Kant acknowledges that sympathetic feelings may produce benevolent actions that are not motivated by self-interest or vanity. On the other hand, although such actions are in conformity with duty (*pflichtmäßig*), and so deserve praise and encouragement, they lack the moral value that deserves genuine esteem (*Hochschätzung*) because they are not done from duty (*aus Pflicht*) (*G*, 4: 398). Rather, a naturally sympathetic person has an immediate inclination to perform such actions. An action is here considered to be less morally praiseworthy because the performance of it too closely aligns with what one already desires to do in consequence of one's character or other circumstances that one has not freely chosen and for which one cannot, therefore, claim any moral credit. In contrast, a person who lacks a sympathetic disposition but, nevertheless, acts benevolently must be assumed to act from a sense of duty alone, for there is no pre-existing inclination that can adequately explain why he or she acts in this way.

Acting from duty consists in adopting a rule of action that by virtue of its universality and unconditionality possesses the status of a law. In order to determine whether a specific rule of action possesses this status, it must be evaluated according to the higher-order principle of lawfulness as such. The identification of this principle requires abstracting from all inclinations and thus the ends towards which they are directed. All that then remains is the form of law, which finds practical expression in the following demand: '*I ought never to act except in such a way that I could also will that my maxim should become a universal law*' (*G*, 4: 402). A maxim is a rule of action that a rational agent proposes to itself. One ought, therefore, to seek to universalise such a rule of action by asking whether it could be adopted and followed by all rational agents without generating a contradiction. This is not to say that no type of feeling is involved. Indeed, Kant associates a specific type of feeling with moral duty. This feeling is respect (*Achtung*) for the moral law. Yet this feeling does not simply happen to us. Rather, it is a feeling that results from a certain type of self-activity, in that it is produced in us by moral demands on us that we ourselves must identify and seek to fulfil.

We can see that Kant has reasons for wanting to downplay the moral value of compassion. Although compassion provides an inclination to help others that may result in morally desirable outcomes which are in conformity with such duties as benevolence, so that actions motivated by compassion may be difficult to distinguish from actions that are motivated by duty alone, it threatens to make moral actions conditional on the existence of this emotion in a person in a particular situation in which moral demands are made on him or her. Nor is there any guarantee that actions motivated by compassion will in fact correspond to that which duty demands, for compassion does not provide an independent standard that shows that what it prompts one to do or to refrain from doing is a genuine moral duty.⁶ Moreover, one cannot be commanded to have compassionate feelings, whereas Kant's concept of duty entails that an action is morally commanded or forbidden irrespective of

whether a person is or is not inclined to perform it. For such feelings are something over which a human being ultimately lacks decisive control, even if they can be cultivated in and by individuals, so that one can, in this respect, speak of duties in relation to them. Kant accordingly distinguishes between ‘pathological’ love, which cannot be commanded, and ‘practical’ love, which concerns the duty to act benevolently, and he contrasts a willing based on principles of action with a ‘melting sympathy’ that is reducible to feeling (G, 4: 399). None of this means, however, that compassion has *no* moral value. Indeed, if it had no such value, how could Kant claim that actions motivated by it may deserve praise and encouragement? Compassion is not, therefore, by its nature incompatible with moral duty.

If compassion can be accorded some moral value, if not a supreme moral value, the question of how it can be accommodated within Kant’s theory of moral agency arises, as well as the question of why it must be accommodated within this theory. Certain claims that Kant makes concerning a propensity (*Hang*) to moral evil which is present in human nature help to answer both questions. This propensity to moral evil concerns the adoption of a maxim that diverges from the moral law, hence the following definition of radical evil: “‘The human being is *evil*,” cannot mean anything else than that he is conscious of the moral law and yet has incorporated into his maxim the (occasional) deviation from it’ (R, 6: 32).

This definition of radical evil can be explained in roughly the following way. Human beings are subject to unconditional moral demands whose source is reason. These demands are derivable from a single moral law or principle that requires the universalisation of any maxim, that is, rule of action, which one proposes to adopt and to follow. Radical evil concerns the choice of the higher-order maxim to obey this law or principle only when doing so does not conflict with the demands of self-love, whereas one ought to do what the moral law commands or forbids simply because it commands or forbids it. The moral law is therefore not completely absent from an agent’s maxim. This maxim is, nevertheless, a rule of action that dictates making exceptions to that law which ought to have been unconditionally adopted as one’s highest principle of action, that is, whenever that which this law commands or forbids is not sufficiently aligned with incentives whose source is self-love. As Kant himself puts it:

It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims. He indeed incorporates the moral law into those maxims, together with the law of self-love; since, however, he realizes that the two cannot stand on an equal footing, but one must be subordinated to the other as its supreme condition, he makes the incentives of self-love and their inclinations the condition of compliance with the moral law – whereas it is this latter that, as *the supreme condition* of the satisfaction of the former, should have been incorporated into the universal maxim of the power of choice as the sole incentive. (R, 6: 36)

Kant explains this failure to will the moral law by subordinating the incentive provided by respect for the moral law to incentives of self-love in terms of the idea of an evil heart, concerning which he identifies three stages. First, there is ‘the general weakness of the human heart’ or ‘the *frailty* of human nature’ that consists in the inability do what is morally good even when one has adopted the good, in the form of

the moral law, as one's maxim, because the corresponding incentive is always weaker in relation to another incentive whose source is an inclination that conflicts with that which the moral law commands or forbids. Second, there is the 'impurity' of the human heart that consists in the need for incentives other than the moral law because the incentive provided by the good is not sufficient to motivate the performance of actions commanded by this law or the omission of actions forbidden by it. One does not, therefore, act 'purely from duty' (*rein aus Pflicht*). Third, there is the 'depravity' or 'corruption' of the human heart that consists in the propensity to choose to subordinate the incentive provided by the moral law to other incentives, thereby reversing the moral order which requires subordinating all other incentives to the only genuine moral incentive (R, 6: 29–30).

It is significant in relation to the question of the moral value of compassion that although the corruption of the human heart is incompatible with genuine moral agency because it consists in a 'reversal of incentives' that is 'contrary to the moral order' (R, 6: 36), the same does not necessarily apply to the impurity of the human heart. For the need for an additional incentive does not entail that the incentive provided by the moral law is reduced to a subordinate one. Rather, one could act primarily from duty if not *purely* from it, so that the 'moral order' of incentives is preserved. This appears possible especially when the additional incentive tends to result in actions that deserve moral praise and encouragement, as Kant suggests is true of inclinations produced within a human being by compassionate feelings. From this we can see how it is possible to accommodate compassion within Kant's theory of genuine moral agency, albeit as a concession to a specific weakness of human nature. The question then becomes how we can reliably tell whether the incentive provided by the moral law maintains its primacy in relation to an additional incentive of this kind. This difficulty helps to explain why, in the following passage, Kant groups compassion together with inclinations whose non-moral, or even immoral, character is more evident:

[W]henever incentives other than the law itself (e.g. ambition, self-love in general, yes, even a kindly instinct such as compassion [*Mitleiden*]) are necessary to determine the power of choice to *lawful* actions, it is purely accidental that these actions agree with the law, for the incentives might equally well incite its violation. The maxim, by the goodness of which all the moral worth of the person must be assessed, is therefore still contrary to law, and the human being, despite all his good actions, is nevertheless evil. (R, 6: 30–1; translation modified)

Yet, is the inclusion of compassion among these other inclinations, not only on the grounds that it is an unreliable guide to whether a proposed rule of action is consistent with the moral law but also on the grounds that it introduces an incentive that may not be subordinate to the incentive of respect for this law, truly justified? Does Kant not thereby imply that the person who is guided by compassion is just as evil as the person who is guided only by self-love simply because there is no guarantee that the relevant type of inclination will motivate genuinely moral actions?

As we shall see in the next section, Rousseau's theory of the moral value of compassion can be used to show how Kant could grant compassion a similar value,

thereby undermining the claim that it is, when judged from a moral perspective, essentially no better than other more obviously non-moral or immoral incentives. This corresponds to how Kant himself suggests that compassion provides an incentive that tends to be compatible with the demands of morality, so that it is not accurate to describe any agreement between this additional incentive and genuine moral actions as 'purely accidental'.

I shall begin with a passage from Kant's *Anthropology from a Pragmatic Point of View* and then explain the significance of this passage in relation to certain claims that Rousseau makes in connection with the role that pity plays in a human being's moral development. This will enable us to see how Kant could grant compassion a moral value that goes beyond the conditional value it possesses as an appropriate supplementary inclination that remains subordinate to the incentive provided by respect for the moral law. Compassion is valuable also because it paves the way for a human being's entry into a moral world in which the truly moral incentive of respect can begin to make its presence felt in him or her.

3. Compassion and moral development

The relevant passage from Kant's *Anthropology from a Pragmatic Point of View* reads as follows:

The principle of *apathy* – namely, that the wise man must never be in a state of affect, not even in that of compassion with the misfortune of his best friend, is an entirely correct and sublime moral principle of the Stoic school; for affect makes us (more or less) blind. – Nevertheless, the wisdom of nature has planted in us the predisposition to compassion in order to handle the reins *provisionally*, until reason has achieved the necessary strength; that is to say, for the purpose of enlivening us, nature has added the incentive of pathological (sensible) impulse to the moral incentives for the good, as a temporary surrogate of reason. (*Anth*, 7: 253)

The first sentence of this passage is compatible with a negative assessment of the moral value of compassion: it is an affective state that provides an inadequate guide to moral action because there is no reliable way of telling whether this emotion is directed towards the right objects in the right way. The second sentence, however, indicates a more positive assessment of the moral value of compassion: it is a natural feeling that provides an incentive to act morally in the absence of a sufficiently strong rational incentive to do so.⁷ The need for this additional incentive can be explained in terms of how reason is insufficiently developed in a person, so that a 'surrogate of reason' is needed. If reason were sufficiently developed, a purely rational incentive would be available because this person would then be in a position to become conscious of the moral law and experience respect for it, though even then, the impurity of the human will would mean that this incentive is unlikely to be by itself sufficient to motivate moral action.

If it is to perform the function of a 'surrogate of reason', compassion must reliably track the demands of morality, making it different from incentives that tend to

motivate actions that are either contrary to these demands or are morally indifferent because they may be in conformity with the demands of morality without there being anything moral about them considered in themselves. In this respect, the passage quoted above is compatible with the following claims: (1) the claim that compassion is natural to human beings and (2) the claim that it may guide human beings to do what is morally right and to avoid doing what is morally wrong independently of reason. Here is the place to introduce Rousseau, not only because he endorses both claims, but also because he assigns to compassion a role in an individual's moral development prior to the emergence of a distinctively moral form of reasoning that implies that this emotion is a condition of this type of reasoning, in that it produces an awareness of the distinctive moral status of the kind of being that is an object of moral duties.

In the *Discourse on the Origin and Foundations of Inequality Among Men*, Rousseau identifies pity with 'a natural repugnance to seeing any sentient Being, and especially any being like ourselves, perish or suffer' (OC, 3: 126; DI, 127). This repugnance is explained in terms of a natural sentiment that disposes human beings to experience the suffering of other sentient beings as somehow painful for themselves as well. This aversion to the suffering of others manifests itself in the form of pity, which is natural in the sense of being a fundamental element of human nature that expresses itself instinctively, leading Rousseau to describe the sentiment of pity as a 'pure movement of Nature' and to speak of 'the force of natural pity' (OC, 3: 155; DI, 152). A human being 'is restrained by Natural pity from doing anyone harm' (OC, 3: 170; DI, 166). Yet pity does something more than prevent human beings from harming one another, for it also 'carries us without reflection to the assistance of those we see suffer' and it is pity that 'in the state of Nature, takes the place of Laws, morals, and virtue, with the advantage that no one is tempted to disobey its gentle voice' (OC, 3: 156; DI, 154). This implies that pity equally motivates human beings to act benevolently in relation to one another prior to any rational reflection. Indeed, since these claims are made in connection with primitive human beings whose ability to reason remains merely latent because they lack the opportunity to exercise the relevant faculties, Rousseau commits himself to the claim that pity is an incentive to moral action that operates independently of reason.

Despite its naturalness, pity forms an object of education in the later *Emile*. Although this suggests a tension between Rousseau's earlier and later accounts of pity because pity no longer appears to be an instinctive response to the suffering of others, this tension can be resolved by viewing pity to be a natural, but merely latent, sentiment that must be awakened at the appropriate time, that is, when required by a natural order in accordance with which the faculties of a human being who is destined to enter society should be exercised and developed.⁸ The way in which pity remains a natural sentiment is shown by the context in which Rousseau introduces it in *Emile*.

The early stages of education are concerned with another natural sentiment. This is self-love (*amour de soi*), which Rousseau identifies with the desire to preserve one's own life and the desire for basic physical and psychological well-being (OC, 3: 125–6; DI, 127). Education here aims at '[t]he internal development of our faculties and our organs' (OC, 4: 247; E, 162) and the production of relevant theoretical and practical knowledge of external objects by means of a correctly ordered, and thus properly regulated, experience of them. This involves one human being teaching another

human being how to make use of his or her physical and mental powers, to comprehend the workings of his or her natural and human environments, and to employ things encountered within them for his or her own benefit. The next stage concerns the establishment of appropriate emotional and psychological ties with other human beings. This requires that the relations between the pupil and others be of the right kind, that is, they must be genuine moral relations. An extension of the pupil's sensibility is a condition of moral relations at the same time as such relations are conditions of becoming truly human: 'It is only when it begins to extend outside of himself that it takes on, first, the sentiments and, then, the notions of good and evil which truly constitute him as a man and an integral part of his species' (OC, 4: 501; E, 371). Rousseau introduces pity in connection with this extension of the pupil's sensibility. The following claim is here a key one: 'The source of all the passions is sensibility; imagination determines their bent [*leur pente*]' (OC, 4: 501; E, 370).

A passion provides an incentive to action. The object of a passion, that is, that towards which it is directed, varies according to the influence exercised upon this passion by the imagination. Therefore, by constraining the imagination, it is possible to ensure that a powerful affect is directed at the right type of object in the right way. For Rousseau, at this stage it is sentiment rather than reason that must constrain the imagination. Pity is a natural sentiment that can perform this function because it is an affect directed at a specific object, namely, human beings who suffer and suffering humanity more generally, and it depends on imaginatively identifying oneself with the suffering of others. Thus the imagination is employed in such a way that it is directed by a specific sentiment and constrained by the specific object of this sentiment. The general impulse to extend one's influence beyond the boundaries of one's own self, which can express itself in other, undesirable ways, such as in the desire to dominate others, is thereby channelled in a moral direction. To ensure that this happens, the pupil must be presented with objects that give this impulse to expand the right direction:

To excite and nourish this nascent sensibility, to guide it or follow it in its natural inclination [*sa pente naturelle*], what is there to do other than to offer the young man objects on which the expansive force of his heart can act – objects which swell the heart, which extend it to other beings, which make it find itself everywhere outside of itself – and carefully to keep away those which contract and concentrate the heart and tighten the spring of the human I [*moi humain*]. (OC, 4: 506; E, 374–5)

Given that pity is the relevant sentiment, the objects to be presented to the pupil are examples of suffering to which he himself, like any other human being, is vulnerable.

From this brief account of the role of pity, we can already begin to see how it involves making human beings responsive to moral demands prior to any consciousness of rational moral norms. This function provides us with a way of understanding Kant's claim in the passage from *Anthropology from a Pragmatic Point of View* quoted above that 'the wisdom of nature has planted in us the predisposition to compassion in order to handle the reins *provisionally*, until reason has achieved the necessary strength'. For although the relevant passage may appear to claim only that the additional but subordinate incentive provided by compassion is necessary to

ensure that one acts morally, there are now two related reasons for thinking that compassion, or pity, is essential to an individual's moral development prior to any consciousness of rational moral norms and the feeling of respect for the moral law that accompanies this consciousness.

First, even if compassion is a natural sentiment for Kant, as pity is for Rousseau, this sentiment needs to be cultivated and properly directed. Second, Kant implies that compassion is already a generally reliable guide to moral action, for if it were not one, it could not fulfil the function of providing an additional incentive to moral actions rather than producing morally inappropriate or indifferent ones, a function that requires reliably tracking the demands of morality prior to reflection on these demands and knowledge of their rational grounds. I shall now show that Rousseau provides an account of how compassion performs the relevant function during the specific phase of moral development in which the transition to the stage of moral reasoning occurs by making the pupil conscious of, and by attuning him to, the type of moral status that is fundamental to Kant's moral theory. This requires looking at what the transition to the stage of moral reasoning involves rather than considering the specific proposals designed to bring about this transition.

Before the sentiment of pity has been awakened in the pupil, his relations to other human beings are of the kind indicated by Rousseau's claim that '[t]he child raised according to his age is alone. He knows no attachments other than those of habit. He loves his sister as he loves his watch, and his friend as his dog' (*OC*, 4: 500; *E*, 370). This claim implies that the pupil views other human beings as things or, at most, as animals in relation to which he experiences an ill-defined emotional attachment. It would not be true to say that the pupil is completely indifferent to other human beings, for they possess an instrumental value for him, with even the pupil's emotional attachment to his sister or his friend being explicable in terms of how the presence of others contributes to his sense of well-being. In this respect, the presence of others matters to the pupil only in so far as it contributes to the satisfaction of desires whose ultimate source is self-love. This is to be expected, given that self-love has so far formed the object of education.

The role of compassion in bringing about a moral change in how the pupil relates to others suggests that experiencing the sentiment of pity leads the pupil to *care about* other human beings and to develop an interest in how well or how badly *their* lives are going. Pity does this by shifting attention away from himself and how others are useful to him to their inner lives and how these lives have meaning and value *for them*. For pitying others not only involves experiencing certain feelings by imagining their suffering and experiencing in one's own person feelings that are typically occasioned by this way of sharing the feelings of others, but also prompts one to ask oneself what the causes of this suffering are. Since the pupil knows from his own experience that the inability to satisfy his desires is the source of suffering in his own person, he must recognise that other human beings have desires as soon as he begins to acknowledge that they suffer. This leads him to ask himself which specific desires explain this suffering.

In this way, pitying others is bound up with a type of rational reflection that is restricted to drawing inferences and making judgements based on beliefs that derive from one's own experiences and from observing the behaviour of others. The desires whose non-fulfilment explains another human being's suffering are related to ends

that these human beings must be thought to value. Thus the process whereby the pupil's attention is directed away from himself and what matters to him to the inner lives of others and what *they* value and what gives *their* lives meaning leads him to view other human beings as agents who possess the capacity to form and adopt their own ends, as opposed to having ends imposed upon them by him in such a way that they are reduced to mere means. How does this help to explain the claim that Kant can be thought to grant compassion an important role in an individual's moral development? This question can be answered with reference to Kant's distinction between a 'market price' that concerns a merely relative form of value and the dignity that derives from an 'inner' worth (G, 4: 434–5).

In so far as things have a price, they can be treated as equivalents that possess a purely instrumental value. This implies the existence of an independent measure of value that determines to what extent qualitatively different things have the same or a different quantitative value. This type of value is relative in two senses: (1) the value of each thing is relative to the value of other equally measurable things, that is, it is of more, less, or equal value than they are, and (2) the value of each thing depends on its relation to something other than itself, such as how it promises to satisfy a desire that someone already has. (2) corresponds to how, in *Emile*, before the sentiment of pity is awakened in him, the pupil values other human beings only because they are useful to him. (1) corresponds to how, in the pupil's eyes, utility, as determined by a person's or a thing's relation to his own needs and desires, provides the only measure of value, with the specific value of other human beings depending on the extent to which they are useful to the pupil as means to his own ends. The concepts of dignity and inner worth, in contrast, signal the inappropriateness of applying a common measure of value, such as that of utility, which determines the value of one human being relative to other human beings. From this it follows that the value of each human being ought not to be regarded as depending on, and thus reducible to, how useful he or she is to others or to society. Kant correspondingly attributes to all rational beings a status which demands that they never be treated merely as means (G, 4: 437–8).

This invites the following question: what is the source of this status and the moral command to respect other persons that it entails? Kant's answer to this question is that the relevant type of value derives from a rational being's capacity to give itself universal laws and how such a being equally stands under such laws (G, 4: 433, 438). This implies that human beings can be subsumed under a general category after all, namely, that of a rational being capable of giving itself universal laws and, as a being of the relevant general type, standing under these laws. This category itself forbids, however, reducing the value of each human being to a merely relative, instrumental one. Yet we may then ask how each human being can come to view other human beings as beings with this dignified status.

It is precisely here that compassion can serve as 'a temporary surrogate of reason' by playing a crucial role in an individual's moral development, for, as we have seen, it is through experiencing compassion in relation to the suffering of other human beings that one becomes aware of the inner lives of others and the meaning and value that their lives and ends have for them. This constitutes a key transitional stage in an individual's moral development, in that others are no longer viewed purely in terms of their usefulness to oneself. Rather, their capacity to set and value their own ends is recognised. These ends may or may not be compatible with one's own ends. If they are

not compatible with one's own ends, then seeking to make others act in conformity with one's own ends is likely to make them suffer in some way, whereas it is precisely this outcome that compassion for others will make one want to avoid.

Compassion can here be viewed as a natural sentiment that leads one human being to treat other human beings in ways that respect the distinction between an instrumental market value and a dignity whose source is the status that rational beings possess as ends in themselves, though even before this difference in status has become an object of reflection and the rational grounds of it have been discovered. In this way, compassion provides an incentive that tracks, in a sufficiently reliable way, certain fundamental moral demands, making it different from incentives that tend to result in actions that are incompatible with these demands or have no recognisable connection with them. Compassion is an emotion that is eminently suited to perform this function in each individual's moral development because of its specific object, which is the suffering of others. For unlike the form of sympathy which consists in feelings of pleasure caused by another person's joy, compassion directly concerns duties that require helping others because there is some way in which their needs are unsatisfied or their ends remain unfulfilled. Another person's joy, in contrast, can be explained in terms of how his or her ends or expectations have been fulfilled.

This would be only a first stage, however, because the dignity that human beings possess is, for Kant, not simply a matter of having ends. Rather, it is a matter of adopting ends that accord with the idea of universal law, that is, ends that can be present in freely adopted universalisable maxims. The access to the inner lives of others made possible by compassion is, nevertheless, essential because it enables one human being to become aware of how other human beings possess the same capacity to set their own ends and to value them, thereby paving the way for evaluation of the morality of these ends. With respect to this next stage, it can be asked whether compassion will eventually be replaced by the guide to moral action provided by reason alone, so that although compassion plays a vital role in an individual's moral development in the way described above, it can be dispensed with once the higher standpoint of rational reflection has been attained.

Compassion would even then have that value which derives from how it directs us to the volitional lives of others in such a way as to predispose us to recognise the distinction between a purely instrumental form of value and the intrinsic value that each human being possesses, thereby making us receptive to moral demands.⁹ In the absence of this affective element of moral experience, there is the danger of adopting the purely objective standpoint of 'a Martian onlooker' which Martha Nussbaum associates with the anti-pity tradition to which she claims Kant belongs (Nussbaum 1996: 47), or that of Rousseau's pupil in relation to his sister and his friend before the sentiment of pity has been awakened in him. This moral value of compassion is independent of its value as an additional but subordinate incentive, for even a person who did not require such an incentive would not only have to become sufficiently attuned to the moral status of others but also remain attuned to it, given the possibility of relapsing into a state in which one tends to view others merely as means to one's own ends. The position that reason alone can be sufficient is, nevertheless, suggested by the following passage from the *Critique of Practical Reason*, even if this position does not exclude the need for a prior stage of moral development in which compassion serves as a temporary substitute for reason:

Even an inclination to what conforms with duty (e.g., to beneficence) can indeed greatly facilitate the effectiveness of *moral* maxims but cannot produce any. For in these everything must be directed to the representation of the law as determining ground if the action is to contain not merely *legality* but also *morality*. Inclination is blind and servile, whether it is kindly or not; and when morality is in question, reason must not play the part of mere guardian to inclination but, disregarding it altogether, must attend solely to its own interest as pure practical reason. Even this feeling of compassion and tender sympathy, if it precedes consideration of what is duty and becomes the determining ground, is itself burdensome to right-thinking persons, brings their considered maxims into confusion, and produces the wish to be freed from them and subject to lawgiving reason alone. (CPrR, 5: 118)

On the one hand, this passage is consistent with how compassion may function as an incentive in addition to the feeling of respect for the moral law, provided the genuine moral order of incentives is not reversed. On the other hand, we have also seen how it then becomes difficult to judge whether respect for the moral law remains the determining ground of the will. This helps to explain Kant's claim that compassion is a source of confusion.

If this is why compassion is burdensome, then the removal of the burden would consist in replacing it with reason alone. Such purity of the will may be extremely difficult, and perhaps impossible, for human beings to achieve. Yet Kant's reduction of compassion to an additional but subordinate incentive, together with the way in which he associates the need for an incentive of this kind with the impurity of the human will, implies that moral purity is an ideal to which any 'right-thinking' person will aspire, and with which he or she is already disposed to identify himself or herself because it is expressive of his or her status as a free and rational being capable of subjecting itself to universal laws that follow from the nature of such a being. Kant must, therefore, remain suspicious of compassion. Are we not then confronted with a fundamental difference between Kant's attitude towards compassion and Rousseau's attitude towards it? Although it is not my intention to argue that there is no significant divergence here, I shall argue in the next section that the difference in question appears less significant if we keep in view the human standpoint emphasised by Rousseau.

4. Compassion and being human

One way of minimising the difference in question would involve interpreting Rousseau's position in such a way as to bring it closer to Kant's by showing how the former anticipates the latter's understanding of the connection between reason and morality and the accompanying devaluation of compassion. An approach of this kind is suggested when it is claimed that for Rousseau pity lacks sufficient generality and fails to provide a reliable guide to moral action. It must, therefore, be guided by abstract ideas of reason, especially that of justice (Neuhouser 2008: 222–3). This is consistent with how Rousseau concedes that the pupil's initial affective bonds with other human beings will be restricted to those individuals with whom he is most familiar and for whom he already experiences a need: 'In directing his nascent

sensibility to his species, do not believe that it will at the outset embrace all men [*tous les hommes*], and that the word *humankind* [*genre humain*] will signify anything to him' (OC, 4: 520; E, 386–7; translation modified). Moreover, Rousseau suggests that reason must replace sentiment when he claims that '[r]eason alone teaches us to know good and evil [*le bien et le mal*]' (OC, 4: 288; E, 196; translation modified). A closer look at these claims, however, reveals the limited extent to which they can be interpreted to mean that reason ought to replace sentiment in the way suggested by the passage from the *Critique of Practical Reason* quoted above.

The claim that only reason provides knowledge of what is morally good and morally evil is followed by the further claim that '[b]efore the age of reason we do good and evil without knowing it, and there is no morality in our actions' (OC, 4: 288; E, 196; translation modified). Rousseau here distinguishes between *doing* what is morally right or morally wrong, on the one hand, and *knowing* that one is doing what is morally right or morally wrong, on the other. Although the latter requires knowledge of the relevant moral concepts, the former does not depend on such knowledge. The actions lack morality, then, only in the sense that they are not done with any explicit knowledge of the moral duties to which they conform.

Doing and knowing are independent of each other at the relevant stage of education because the pupil is motivated by the natural sentiment of pity to help others and to avoid harming them before he has begun to reflect on what he ought and ought not to do. This does not mean that activities associated with reason, such as inference and judgement, are not at work. Indeed, near the end of the third book of *Emile*, Rousseau implies that these activities are presupposed by that stage during which moral relations form the object of education, for he states that '[w]e have made an active and thinking being. It remains for us, in order to complete the man, only to make a loving and feeling being – that is to say, to perfect reason by sentiment' (OC, 4: 481; E, 353). The issue is whether reason, as the source of knowledge of moral concepts and principles, is strictly necessary for actions to be judged morally good or morally bad and whether reason should seek to operate independently of pity. This brings me to the claim that pity is insufficiently general.

Although Rousseau speaks of how the pupil can generalise 'his individual notions under the abstract idea of humanity', his characterisation of this process of generalisation is instructive, for it emerges from 'many reflections on his own sentiments and on those he observes in others', and it involves adding to his 'particular' affections 'those which can make him identify with his species' (OC, 4: 520; E, 387). For reasons that have already been set out, pity is key to establishing this affective identification with other human beings. Moreover, Rousseau mentions that if it had been the right place, he would have shown in connection with the pupil's entry into 'the moral order' how '*justice and goodness* are not merely abstract words – pure moral beings formed by the understanding – but are true affections of the soul enlightened [*éclairée*] by reason, are hence only an ordered development of our primitive affections' (OC, 4: 522–3; E, 388–9).

From such claims it is evident that, for Rousseau, moral concepts have an affective basis that reason can clarify and build upon but never completely transcend. This means that even if pity is a sentiment that must ultimately be guided by reason in the form of abstract moral ideas, it remains the affective basis of these ideas themselves. Reason cannot, therefore, operate as an independent moral authority. Moreover,

Rousseau assumes that pity is a sufficiently reliable guide to morally right actions when he explains these moral ideas in terms of a process through which this sentiment is generalised. Rousseau's own reason for generalising pity to the whole human race is to prevent it from 'degenerating into weakness' and to ensure that 'one yields to it only insofar as it accords with justice, because of all the virtues justice is the one that contributes most to the common good of men' (*OC*, 4: 548; *E*, 409). Although the idea of justice may here serve to correct what the person overcome by pity might otherwise have unreflectively been led to do, this could mean only that this idea stabilises, as it were, the sentiment of pity by providing it with some general rules that express what it itself already prompts human beings to do or to refrain from doing.

Kant's reduction of compassion to a concession to the impurity of the human will indicates that this emotion can be considered inessential when viewed from the ideal moral standpoint suggested by the idea of a 'holy will'. The concept of morality (*Sittlichkeit*) is valid not only for human beings and under contingent conditions. Rather, it must be considered absolutely and necessarily valid for all rational beings (*G*, 4: 408). The concept of duty, in contrast, is restricted to a good will subject to 'certain subjective limitations and hindrances' (*G*, 4: 397). The concept of duty would therefore not apply to a will that is free from these subjective limitations and hindrances at the same time as it is, like any rational will, subject to the concept of morality and all that directly follows from it: 'A perfectly good will would . . . equally stand under objective laws (of the good), but it could not on this account be represented as *necessitated* to actions in conformity with law since of itself, by its subjective constitution, it can be determined only through the representation of the good' (*G*, 4: 414). This is a will that would not be tempted to exempt itself from the moral law whenever self-love dictates this. It is, therefore, not a will that would require the additional but subordinate incentive provided by compassion. As an ideal of pure practical reason, this idea of a holy will entails that human beings should seek to become morally pure by freeing themselves of any subjective limitations or hindrances, among which we may include the incentive provided by compassion, which is a concession to human moral weakness.

Rousseau can be seen to oppose the kind of moral ideal represented by Kant's idea of a holy will when he stresses the existence of objective 'limitations and hindrances' in the form of needs and vulnerabilities to which human beings are subject, and which are the ultimate sources of the actual and potential suffering that unites them as equals:

[W]e are attached to our fellows [*nos semblables*] less by the sentiment of their pleasures than by the sentiment of their pains, for we see far better in the latter the identity of our natures with theirs and the guarantees of their attachment to us. If our common needs unite us by interest, our common miseries unite us by affection. (*OC*, 4: 503; *E*, 373)

Human beings are not materially, emotionally, and psychologically self-sufficient beings. Their dependence on one another increases their vulnerability to suffering because there is no guarantee that others will respond to one's needs and desires in the ways that one wants them to. Since suffering results from some kind of failure to

meet one's needs or to satisfy one's desires, including the desire to avoid pain, Rousseau contrasts the 'frail happiness' of which human beings are capable with the secure and complete happiness of a solitary and self-sufficient being. This is the 'absolute happiness' enjoyed by God. Yet the idea of a being that enjoys this form of happiness is so alien to what it means to be human that human beings could not conceive of this solitary and self-sufficient being's condition as a truly happy one. Rather, relationships with other human beings are integral to any human conception of happiness. A human being must therefore view the solitary and self-sufficient nature of this being as a source of unhappiness (*OC*, 4: 503; *E*, 372).¹⁰ Analogously, if we replace the idea of a divine form of happiness with the moral self-sufficiency of Kant's holy will, one could argue that human beings must struggle to form any genuine idea of this holy will, which must therefore remain alien to the moral lives of human beings.

As we have seen, it is precisely in connection with the moral life of a human being who is far from attaining the standpoint of a holy will that Kant assigns an important role to compassion. This is not only because compassion provides the right kind of additional but subordinate incentive that the impurity of the human will makes necessary, but also because compassion plays an essential role in an individual's moral development by attuning him or her to the moral status of others, thereby making him or her responsive to the demands that follow from this status, even if reason is the ultimate source of this status and these demands. Moreover, human beings are confronted with obstacles when it comes to remaining attuned to the moral status of others, especially the temptation to reduce other human beings to means to their own ends, which is likely to be strong because of factors such as the need to survive and the desire to prosper in a competitive society. The role of compassion in a human being's moral development cannot, therefore, be reduced to a 'surrogate of reason' that can be discarded once the stage of moral reasoning has been attained. Rather, even when this stage has been attained, compassion will continue to perform the important function of ensuring that human beings remain attuned to the moral status of others and responsive to the demands that follow from this status. Thus, from the human standpoint, compassion is an emotion that possesses considerable moral value for Kant as well as for Rousseau, even if the latter may grant it a greater and more independent moral value than the former does.

Notes

1 I use the following abbreviations: *Anth* = *Anthropology from a Pragmatic Point of View* (in Kant 2007); *CpR* = *Critique of Practical Reason* (in Kant 1996); *G* = *Groundwork of the Metaphysics of Morals* (in Kant 1996); *MM* = *The Metaphysics of Morals* (in Kant 1996); *R* = *Religion Within the Boundaries of Mere Reason* (Kant 1998); *ROFBS* = 'Remarks in the Observations on the Feeling of the Beautiful and the Sublime' (in Kant 2011). The pagination is that of the Akademie edition of Kant's writings.

2 Among the earliest and most comprehensive studies in this vein is Ernst Cassirer's essay 'Rousseau and Kant' in Cassirer (1945). For an overview of some common themes that have been identified and an attempt to extend the range of them to include Rousseau's epistemology and his theory of cognitive development that appeals to the educational project set out in *Emile*, see Hanley (2017). I intend to show with reference to the theory of moral development found in *Emile* and the central role played by pity in this theory that there is another common theme, if only implicitly so.

3 I use the term ‘pity’ in connection with Rousseau’s writings because it is the most natural way to translate the French word *pitié* even if this word could be translated as ‘compassion’. In addition to compassion, which is the translation of the German word *Mitleid*, Kant refers to ‘an insulting kind of beneficence, since it expresses the kind of benevolence one has toward someone unworthy, called *pity* [*Barmherzigkeit*]’ (*MM*, 6: 457). This type of pity is accompanied by a sense of superiority in relation to the person whose suffering or unhappy condition forms the object of it. It can therefore be thought to belong to a ‘proto-Nietzschean’ critique of pity. See Sorensen (2018: esp. 215, 217). Yet the use of the English word ‘pity’ to translate the German word *Barmherzigkeit* should not lead us to confuse it with what Rousseau means by *pitié*, which broadly corresponds to *Mitleid*. Moreover, the claim that sympathy has a ‘positive valence’ for Kant, which partly concerns its role in the moral development of human beings, whereas compassion functions as a ‘near synonym’ for its ‘negative valence’, is, or so I intend to show, misleading because compassion (or pity), in the sense intended by Rousseau, can be seen to play a crucial role in Kant’s understanding of moral development. For the relevant claims, see Sorensen (2018: 209, 216).

4 I use the following abbreviations for Rousseau’s writings: *DI* = *Discourse on the Origin and Foundations of Inequality Among Men* (Rousseau 1997); *E* = *Emile or On Education* (Rousseau 2010); *OC* = *Œuvres complètes* (Rousseau 1959–1995: cited by volume and page number).

5 For Kant, this purely passive type of relation is exemplified by feelings that communicate themselves from one person to another like some kind of disease and the type of shared affective state (*Mitleidenschaft*) that ‘spreads naturally among human beings living near one another’ (*MM*, 6: 457).

6 For more on this point, see Guyer (2012: 424–5).

7 In her attempt to locate Kant firmly in an anti-pity tradition, Nussbaum refers only to Kant’s acceptance of the Stoic viewpoint. See Nussbaum (1996: 46). Yet we can now see why, despite his praise for this viewpoint, Kant’s endorsement of it is a qualified one.

8 See James (forthcoming).

9 It has been argued that for Kant there is a rational form of sympathy that is not tied to feeling and whose conditional value consists in how it provides ‘moral information’ regarding such factors as the well-being or suffering of others and morally relevant features of the circumstances in which they find themselves. See Hildebrand (2023). Given how feeling is a constitutive feature of it, compassion would be essentially different from a rational form of sympathy that is somehow not tied to feeling. Moreover, the value of compassion is not reducible to how it provides moral information of the type mentioned above. For compassion additionally ensures that we no longer regard others purely as means to our own ends and that we concomitantly recognise that they have ends which they themselves value, the fulfilment of which is essential to their well-being. Since recognition of the moral status of others as ends in themselves explains why their well-being and the specific ends and circumstances bound up with it are morally relevant, compassion must be thought to enjoy a certain primacy in relation to any predominantly cognitive form of sympathy, assuming that Kant endorses some version of the latter. For it is only by becoming attuned to the moral status of others, and thereby becoming conscious of moral demands made on him or her in relation to them, that a person would consider himself or herself obliged to seek morally relevant information about another person’s well-being.

10 Although Kant’s holy will can be thought to be free from any inclinations that are opposed to the moral law and to be self-sufficient in the sense that it stands in no relation of dependency on something other than itself, his account of a person’s attempt to universalise the maxim of not helping others when they are in need acknowledges that human beings, in contrast, are dependent on one another for the satisfaction of their needs, for he speaks of ‘rational beings with needs, united by nature in one dwelling place so that they can help one another’ (*MM*, 6: 453). Against this background, the maxim in question turns out not to be universalisable, thereby establishing the duty to be beneficent.

References

- Cassirer, Ernst (1945) *Rousseau, Kant, Goethe: Two Essays*. Trans. James Gutmann, Paul Oskar Kristeller, and John Herman Randall, Jr. Princeton, NJ: Princeton University Press.
- Guyer, Paul (2012) ‘Schopenhauer, Kant and Compassion’. *Kantian Review*, 17(3), 403–29.

- Hanley, Ryan Patrick (2017) 'Rethinking Kant's Debts to Rousseau'. *Archiv für die Geschichte der Philosophie*, 99(4), 380–404.
- Hildebrand, Carl (2023) 'Feeling, cognition, and the eighteenth-century context of Kantian sympathy'. *British Journal for the History of Philosophy*, 31(5), 974–1004.
- James, David (forthcoming) 'Rousseau on the Natural Goodness of Pity'. In Keith Ansell-Pearson and David James (eds.), *The Empathetic Emotions in the History of Philosophy*. Oxford: Oxford University Press.
- Kant, Immanuel (1996) *Practical Philosophy*. Trans. and ed. Mary J. Gregor. Cambridge: Cambridge University Press.
- (1998) *Religion within the Boundaries of Mere Reason*. Trans. and ed. Allen Wood and George di Giovanni. Cambridge University Press.
- (2007) *Anthropology, History, and Education*. Ed. Robert B. Loudon and Günter Zöller. Cambridge: Cambridge University Press.
- (2011) *Observations on the Feeling of the Beautiful and the Sublime and Other Writings*. Ed. Patrick Frierson and Paul Guyer. Cambridge: Cambridge University Press.
- Neuhouser, Frederick (2008) *Rousseau's Theodicy of Self-Love: Evil, Rationality, and the Drive for Recognition*. Oxford: Oxford University Press.
- Nietzsche, Friedrich (1994) *On the Genealogy of Morality*. Ed. Keith Ansell-Pearson, trans. Carol Diethe. Cambridge: Cambridge University Press.
- Nussbaum, Martha (1996) 'Compassion: The Basic Social Emotion'. *Social Philosophy and Policy*, 13(1), 27–58.
- Rousseau, Jean-Jacques (1997) *Discourse on the Origin and Foundations of Inequality among Men*. In *The Discourses and Other Early Political Writings*. Trans. and ed. Victor Gourevitch. Cambridge: Cambridge University Press.
- (2010) *Emile or On Education*. In *The Collected Writings of Rousseau*, Vol. 13. Trans. and ed. Christopher Kelly and Allan Bloom. Hanover, NH, and London: University Press of New England.
- (1959–1995) *Œuvres complètes*. 5 vols. Ed. Bernard Gagnebin and Marcel Raymond. Paris: Gallimard, Bibliothèque de la Pléiade.
- Schopenhauer, Arthur (2009) *Prize Essay on the Basis of Morals*. In *The Two Fundamental Problems of Ethics*. Trans. and ed. Christopher Janaway. Cambridge: Cambridge University Press.
- Sorensen, Kelly (2018) 'Sympathy, Love, and the Faculty of Feeling'. In Kelly Sorensen and Diane Williamson (eds.), *Kant and the Faculty of Feeling*. Cambridge: Cambridge University Press.